

A Tractable Variant of Cover Time

Agelos Georgakopoulos*

June 29, 2012

Abstract

We introduce a variant of the cover time of a graph, called cover cost, in which the cost of a step is proportional to the number of yet uncovered vertices. It turns out that cover cost is more tractable than cover time; we provide an $O(n^4)$ algorithm for its computation, as well as some explicit formulae. The two values are not very far from each other, and so cover cost might be a useful tool in the study of cover time.

1 Introduction

The cover time $CT_r(G)$ of a graph G from a vertex r is the expected number of steps it takes for random walk on G starting at r to visit all vertices. The cover time of G is defined as $CT(G) := \max_{r \in V(G)} CT_r(G)$. It has been extensively studied in various contexts; applications include the construction of universal traversal sequences [AKL⁺79, Bro90], testing graph connectivity [AKL⁺79, KR95], and protocol testing [MP94]. It has been studied by physicists interested in the fractal structure of the uncovered set of a finite grid; see [DPRZ01] for references and for an interesting relation between the cover time of a finite grid and Brownian motion on Riemannian manifolds.

Many bounds on cover time for specific classes of graphs have been obtained, see e.g. [BW90, Fei95, GWb] and references therein. However, in general it is very hard to obtain exact formulae for cover time, or compute it algorithmically. A question of [AF] that remained open for many years despite several efforts [Mat88, KKL00], and was recently resolved in the affirmative using heavy probabilistic machinery [DLP], is whether there is a deterministic algorithm which approximates $CT(G)$ up to a constant factor in polynomial time.

The following situation suggests a similar concept, which we will call the *cover cost* of G . Suppose that a truck starts at r loaded with some goods to be equally distributed to the other vertices of the graph and performs random walk, leaving the goods corresponding to each vertex the first time it gets there and carrying all remaining goods along. Rather than the expected number of steps it will take to finish this tour (which equals $CT_r(G)$), the truck driver is more interested in the expected total amount of goods that will have to be carried.

*Supported by FWF Grant P-24028-N18.

Definition 1.1. Consider a cover tour of random walk on a graph $G = (V, E)$ starting at vertex r . Define the cost of step i to be $1 - \frac{k}{n}$ where k is the (random) number of vertices visited so far excluding r , and $n := |V| - 1$. Define the cover cost $cc_r(G)$ to be the expected total cost of all steps in the cover tour.

As an example, let S_n be a star with center r and n leaves. The problem of determining $CT_r(S_n)$ is the well known coupon collection problem, and one has $CT_r(S_n) \approx 2n \ln n$ [AF]. It is easy to check that $cc_r(S_n) \approx 2n$: having visited the first k leaves, it will take an expected $\frac{2n}{n-k}$ steps to reach the next unvisited vertex, and almost each such step will cost $\frac{n-k}{n}$; the step from the newly reached leaf back to r will be cheaper by $1/n$, and so $cc_r(S_n) = 2n - 1$.

For reasons that will become clear soon, we also define the (uppercase) *Cover Cost* by $CC_r(G) := ncc_r(G)$ for $n = |V| - 1$. Note that $cc_r(G) < CT_r(G) < CC_r(G)$. There is a further intuitive way of defining Cover Cost. Suppose now that instead of a truck driver in the above game we have an electrician trying to visit all vertices of a graph to do some repair. Define his cost $ECC_r(G)$ to be the sum of the expected waiting times of his clients¹. An easy double-counting argument implies that

$$ECC_r(G) = CC_r(G).$$

Note that, by linearity of expectation, $ECC_r(G) = \sum_{x \in V} H_r(x)$, where $H_r(x)$ is the *hitting time* from r to x , i.e. the expected time for random walk starting at r to reach x . Thus

$$CC_r(G) = \sum_{x \in V} H_r(x). \quad (1)$$

This formula yields a (deterministic) algorithm with running time $O(n^4)$ computing cover cost precisely. We describe this algorithm in Section 4.

In Section 3 we prove

Theorem 1.2. For every graph $G = (V, E)$ and every $r \in V$, we have

$$CC_r(G) = \sum_{x, y \in V} \frac{p_r(x < y)}{p_{xy}}$$

where p_{xy} is the probability that random walk from x will visit y before returning to x , and $p_r(x < y)$ is the probability that random walk from r will visit x before returning to y .

In the case where G is a tree, Theorem 1.2 yields a simpler expression for cover cost involving no probabilistic parameters (see Corollary 3.1), and it implies a surprising connection to the Wiener index proved in [GWa]; see Section 3.

In the case where G is the path P_n of n edges we show that cover cost has the same order of magnitude $O(n^2)$ as cover time. However, the two parameters behave differently when n is fixed and we vary the starting point of our cover tour on P_n . Both $cc_r(P_n)$ and $CT_r(P_n)$ are minimised when r is an endpoint and maximised when r is a midpoint, but the difference between the two extremes behaves very differently: for $CT_r(P_n)$ this difference is also $O(n^2)$, about $CT_r(P_n)/4$, while for $cc_r(P_n)$ the difference is $O(n)$.

¹This parameter was proposed by Peter Winkler (private communication).

Theorem 1.3. *For a path P_n with an even number of edges n , cover cost $cc_r(P_n)$ is minimised when r is an endpoint and maximised when r is the mid-point. Moreover, $cc_{\frac{n}{2}}(P_n) - cc_0(P_n) = n/4$.*

We prove this in Section 2; in fact, we will give an exact formula for $cc_r(P_n)$ for every starting point r .

In Section 5 we propose some problems on cover cost and its relation to cover time.

Throughout this paper, a *random walk* begins at some vertex r of a finite graph G , and when at vertex x , it chooses one of the neighbours of x at random according to the uniform distribution, and moves to that neighbour.

2 Paths

In this section we obtain an explicit formula for the cover cost of a path, from which Theorem 1.3 immediately follows:

Proposition 2.1. *Let P_n be the path on $n + 1$ vertices, indexed by $0, 1, \dots, n$. Then $cc_r(P_n) = \frac{(n+1)(2n+1)}{6} + \frac{r(n-r)}{n}$.*

Proof. We are going to consider the sum of the hitting times $H_r(k)$ from r and apply (1). For $r = 0$ we have the well known formula $H_0(k) = k^2$. For $k > r > 0$, we have $H_0(k) = H_0(r) + H_r(k)$. Combining these two formulas we get $H_r(k) = k^2 - r^2$ for $k > r$, and similarly we get $H_r(k) = (n-k)^2 - (n-r)^2 = (r-k)(2n - (k+r))$ for $k < r$.

Thus we have

$$\begin{aligned} \sum_k H_r(k) &= \sum_{k>r} (k^2 - r^2) + \sum_{k<r} (r-k)(2n - (k+r)) \\ &= \sum_{k>r} (k^2 - r^2) + \sum_{k<r} (k-r)(k+r) + \sum_{k<r} 2n(r-k) \\ &= \sum_{k \neq r} (k^2 - r^2) + \sum_{k<r} 2n(r-k), \end{aligned}$$

and using the formulae for the sum of the first r squares and the first r natural numbers we can rewrite this as

$$\begin{aligned} \sum_k H_r(k) &= \frac{n(n+1)(2n+1)}{6} - r^2 - nr^2 + 2n \frac{(r+1)r}{2} \\ &= \frac{n(n+1)(2n+1)}{6} + r(n-r). \end{aligned}$$

Plugging this into (1) we obtain the desired formula. \square

3 General Formulae for Cover Cost

Let $G = (V, E)$ be a graph on $n + 1$ vertices, and fix $r \in V$. We can express CC_r as the sum of the contribution of each $x \in V$ to CC_r as follows. Start a random

walk particle of ‘charge’ 1 at r , and each time the particle visits a vertex in $V \setminus \{r\}$ for the first time, reduce its charge by $1/n$, letting it continue its random walk with the remaining charge as long as this charge is non-zero. Note that the particle is stopped upon completing a cover tour. For every $x \in V$, let $D(x)$ denote the expected total amount of charge departing from x in this random walk. By the definitions, we have

$$\frac{CC_r(G)}{n} = cc_r(G) = \sum_{x \in V} D(x). \quad (2)$$

Next, we are going to split each $D(x)$ into contributions of each $y \in V$ as follows. Let $V_x^{\dagger y}$ denote the expected number of times that random walk from x will visit x before visiting y for the first time; we also count the starting step as a visit to x , and so $V_x^{\dagger y} > 1$ for every $y \neq x$. Let $p_r(x < y)$ denote the probability for random walk from r to visit x before y .

We claim that

$$D(x) = \frac{1}{n} \sum_{y \in V \setminus \{x\}} p_r(x < y) V_x^{\dagger y}. \quad (3)$$

To see this, think of the initial charge of the particle as having been divided into n ‘quarks’ of charge $1/n$ before beginning the tour, each quark labelled by a distinct vertex at which it is meant to be left. This allows us to write $D(x)$ as the sum of the expected contributions of each quark to D . Linearity of expectation now implies the above formula.

Using the formula for the expected number of repetitions of a Bernoulli trial until the first success, we see that $V_x^{\dagger y} = 1/p_{xy}$ where p_{xy} is the probability that random walk from x will visit y before returning to x . Combining this to the above formulas we get

$$CC_r(G) = \sum_{x, y \in V} \frac{p_r(x < y)}{p_{xy}}. \quad (4)$$

This proves Theorem 1.2.

In the case where G is a tree, the probabilistic parameters in (4) can be replaced by explicit graph-theoretic ones, yielding a pleasant formula for the cover cost. Given vertices r, x, y on a tree, let $x \wedge_r y$ denote the *confluent* of x, y with respect to r , i.e. the vertex of minimal distance from r separating x from y . Recall that $d(x)$ is the degree of x . We have

Corollary 3.1 ([GWA]). *Let T be a tree and $r \in V(T)$. Then*

$$CC_r(T) = \sum_{x, y \in V(T)} d(x \wedge_r y) d(x).$$

This is proved in [GWA], where it is further used to obtain an interesting connection between $CC_r(T)$ and the Wiener index $W(T)$ of T : it is proved that, for every tree T ,

$$\sum_{v \in V(T)} (H_{rv} + d(r, v)) = CC_r(T) + \sum_{v \in V(T)} d(r, v) = 2W(T) := \sum_{x, y \in V(T)} d(x, y).$$

4 Algorithms

Formula (1) allows for an efficient computation of the exact value of cover time: fixing $x \in V$, we can write

$$H_y(x) = 1 + \frac{\sum_{\{z|yz \in E\}} H_z(x)}{d(y)} \quad (5)$$

for every $y \neq x \in V$, since the first step of random walk from y takes us to some neighbour of y . This, and the fact that $H_x(x) = 0$, yields a system of n linear equations with n unknowns, which can be solved in time $O(n^3)$, e.g. by Gaussian elimination. In order to obtain $CC_r(G)$ it suffices, by (1), to solve n such systems, one for each $x \neq r$, and add up the values $H_r(x)$. Thus $CC_r(G)$ can be computed in time $O(n^4)$; in fact, with almost no additional effort we compute simultaneously the Cover Cost for every starting vertex r .²

A further algorithm for the computation of $CC_r(G)$ can be derived from the proof of Theorem 1.2 as follows. For the quantity $D(r)$ of (2) we have, by (3) and the discussion following it, $D(r) = \sum_{y \in V} 1/p_{ry}$. Now each value p_{ry} can be computed by solving a linear system of n equations similar to (5): we have $p_{ry} = \sum_{\{z|rz \in E\}} p_z(y < r)$ and $p_z(y < r) = \sum_{\{z'|zz' \in E\}} p_{z'}(y < r)$ for every $z \neq r, y$, while $p_y(y < r) = 1$. Once $D(r)$ has been computed, all other values $D(x), x \in V \setminus \{r\}$ can be obtained simultaneously by solving a single linear system of n equations: by the definition of $D(x)$ we have

$$D(x) = \sum_{\{z|xz \in E\}} \frac{D(z)}{d(z)}.$$

Thus this method yields a further $O(n^4)$ algorithm for the computation of $CC_r(G)$.

5 Further Problems

The examples in the introduction show that cover cost might, depending on the graph, have the same order of magnitude as cover time or be quite smaller. How does cover cost behave in the extremal cases of ‘fast graphs’, i.e. when $CT = O(n \ln n)$, and ‘slow graphs’, i.e. when $CT(G) = O(n^3)$? (It is known that $n \ln n \lesssim CT(G) \lesssim 4n^3/27$ for every graph G on n vertices [Fei95]).

Problem 5.1. *Is it true that for every $M \in \mathbb{R}^+$ there is $c(M)$ such that for every graph G and $r \in V(G)$, if $CT_r(G) < Mn \ln n$ then $cc_r(G) < c(M)n$, where $n = |V(G)| - 1$?*³

Problem 5.2. *Is it true that for every $M \in \mathbb{R}^+$ there is $c(M)$ such that for every graph G and $r \in V(G)$, if $CT_r(G) > Mn^3$ then $cc_r(G) > c(M)n^3$?*

The extremal graphs for cover time are not known, although a lot of work has been done and graphs that are close to being extremal are known [Fei97]. It would be interesting to find the extremal graphs for cover cost:

²I was made aware of this algorithm by Erol Pekoz.

³This problem arrived after a discussion with Itai Benjamini.

Problem 5.3. Which rooted graph on n vertices minimises $cc_r(G)$? Which maximises it?

It would be very interesting to obtain bounds on $cc_r(G)/CT_r(G)$, for this would allow us to use cover cost in order to obtain bounds on cover time. The following two problems are motivated by this.

Conjecture 5.4. The path on n vertices rooted at an endpoint maximises $cc_r(G)/CT_r(G)$ over all rooted graphs G on n vertices.

Problem 5.5. Is there a graph G for which $cc_r(G)/CT_r(G) < 1/H_{(|V(G)|-1)}$ where $H_n \approx \ln n$ is the n th harmonic number? Which graph on n vertices minimises $cc_r(G)/CT_r(G)$?

Acknowledgements

I am very grateful to Itai Benjamini, Erol Pekoz, and Peter Winkler for their aforementioned suggestions and other valuable discussions.

References

- [AF] D.J. Aldous and J. Fill, *Reversible markov chains and random walks on graphs*, Monograph in preparation. <http://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [AKL⁺79] R. Aleliunas, R. Karp, R. Lipton, L. Lovász, and C. Rackoff, *Random walks, universal traversal sequences, and the complexity of maze traversal*, Proc. 20th IEEE Symp. Found. Computer Science (1979), 218–233.
- [Bro90] A. Broder, *Universal sequences and graph cover times; a short survey*, Sequences (Naples/Positano, 1988) Springer, New York (1990), 109–122.
- [BW90] G. R. Brightwell and P. Winkler, *Extremal cover time for random walks on trees*, J. Graph Theory **14** (1990), no. 5, 547–554.
- [DLP] J. Ding, J.R. Lee, and Y. Peres, *Cover times, blanket times, and majorizing measures*, To appear in Annals of Math.
- [DPRZ01] A. Dembo, Y. Peres, J. Rosen, and O. Zeitouni, *Cover times for brownian motion and random walks in two dimensions*, Acta. Math. **186** (2001), no. 2, 239–270.
- [Fei95] U. Feige, *A tight upper bound on the cover time for random walks on graphs*, Random Structures and Algorithms **6** (1995), no. 1, 51–54.
- [Fei97] ———, *Collecting Coupons on Trees, and the Analysis of Random Walks*, Computational Complexity **6** (1996/1997), 341–356.
- [GWa] A. Georgakopoulos and S. Wagner, *In preparation*.

- [GWb] A. Georgakopoulos and P. Winkler, *New bounds for edge-cover by random walk*, Preprint 2011.
- [KKLV00] J. Kahn, J.H. Kim, L. Lovasz, and V.H. Vu, *The cover time, the blanket time, and the matthews bound*, Proceedings of the 41st Annual Symposium on Foundations of Computer Science, FOCS '00, IEEE Computer Society, 2000, p. 467.
- [KR95] A. Karlin and P. Raghavan, *Random walks and undirected graph connectivity; a survey*, Discrete Probability and Algorithms (Minneapolis 1993) IMA Vol. Math. Appl. Springer, New York, vol. 72, 1995, pp. 95–101.
- [Mat88] P. Matthews, *Covering problems for brownian motion on spheres*, Ann. Probab. **16** (1988), no. 1, 189–199.
- [MP94] M. Mihail and C. Papadimitriou, *On the random walk method for protocol testing*, Computer Aided Verification (Stanford, CA) Lecture Notes on Comp. Sci. 818, 1994, pp. 132–141.